

(19)日本国特許庁 (JP)

(12) 特許公報 (B2)

(11)特許番号

特許第3131432号
(P3131432)

(45)発行日 平成13年1月31日(2001.1.31)

(24)登録日 平成12年11月17日(2000.11.17)

(51)Int.Cl.⁷

G 0 6 F 17/28

識別記号

F I

G 0 6 F 17/28

Q

請求項の数2(全10頁)

(21)出願番号 特願平1-303213

(22)出願日 平成1年11月24日(1989.11.24)

(65)公開番号 特開平3-164975

(43)公開日 平成3年7月16日(1991.7.16)

審査請求日 平成8年11月22日(1996.11.22)

(73)特許権者 99999999

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 伊藤 悅雄

神奈川県川崎市幸区小向東芝町1番地

株式会社東芝総合研究所内

(72)発明者 吉村 裕美子

神奈川県川崎市幸区小向東芝町1番地

株式会社東芝総合研究所内

(72)発明者 長谷部 浩一

神奈川県川崎市幸区小向東芝町1番地

株式会社東芝総合研究所内

(74)代理人 99999999

弁理士 外川 英明 (外1名)

審査官 宮司 卓佳

最終頁に続く

(54)【発明の名称】 機械翻訳方法及び機械翻訳装置

1

(57)【特許請求の範囲】

【請求項1】各種情報を入力するための入力手段と、各種情報を出力するための出力手段と、各種情報を記憶する記憶手段を備えた装置上で、原文情報を翻訳処理して訳文情報を得る機械翻訳処理方法であって、
前記入力手段から入力された強調個所を示す属性データを含む翻訳対象となる原文情報を前記記憶手段に記憶させ、
前記記憶手段から読み出した原文情報に強調個所を示す属性データが含まれているか否かを判定し、
原文情報中に強調個所を示す属性データが含まれていると判定した場合には、翻訳結果である訳文情報を生成する過程で語順に自由度のある部分に対して該属性データが付与されている語句を連続した語順となるように訳文情報を生成し、

10

2

前記生成された訳文情報の対応する個所に強調個所を示す属性データを反映させて前記記憶手段に記憶させ、
前記記憶手段から読み出した訳文情報を前記出力手段に出力させる
ことを特徴とする機械翻訳方法。

【請求項2】強調箇所を示す属性データが含まれた翻訳対象となる原文情報を記憶する原文情報記憶手段と、
前記原文情報記憶手段から読み出した原文情報を強調箇所を示す属性データが含まれているか否かを判定する属性判定手段と、

この属性判定手段で原文情報中に属性データが含まれていると判断した場合には、翻訳結果である訳文情報を生成する過程で語順に自由度のある部分に対して該属性データが付与されている語句を連続した語順となるように訳文を生成すると共に前記属性データを訳文情報の対応

する箇所に反映させる翻訳処理手段とを具備することを特徴とした機械翻訳装置。

【発明の詳細な説明】

【発明の目的】

(産業上の利用分野)

本発明は第1言語のデータを第2言語のデータに翻訳する機械翻訳方法及び機械翻訳装置に関する。

(従来の技術)

先ず、従来の機械翻訳装置の処理について説明する。第1図はその全体ブロック図である。この図によれば、入力部1を介して第1言語のデータを入力すると、翻訳制御部6を介して翻訳部5が辞書記録部4を参照しながら入力した第1言語のデータを各文単位で逐次翻訳することにより第2言語のデータを得、これを再び翻訳制御部6を介して表示部2に表示する。また、入力部1を介して入力された第1言語のデータと翻訳部5により得られた第2言語のデータとを対応づけて記録する文書記録部3を有するものである。

上述した構成によって第1言語のデータを第2言語のデータに翻訳処理する際、従来の機械翻訳装置は、辞書記録部4に概念や構文を与える規則などを記録することにより、第1言語のデータにおいて強調構文や反復等により強調されている箇所に対し、対応する第2言語のデータも強調した形で訳出することが可能である。

しかし、第1言語のデータに下線や斜文字等の文中の他の文字と異なるフォントなどを用いた表記による強調がなされたり、利用者が強調箇所を指定した第1言語のデータを翻訳処理する場合、第1言語のデータで強調された箇所が対応する第2言語のデータには反映されなかった。これは上記した下線や斜文字等の表記情報は、第1言語のデータを第2言語のデータに翻訳処理する際、その対象が言語データのみであるため、その過程で欠落するためである。また、利用者が強調箇所を指定しても、強調構文や反復等でなければ、翻訳結果に反映されることはなく、オペレーターが翻訳処理の後で修正しなければならないという問題点があった。

この他に、第1言語のデータにおいて複数の語句が強調されている場合に、翻訳処理を行なうと、第2言語のデータでは分散されて訳出される場合がある。これは例えば英語のデータと日本語のデータ間の翻訳の場合、それぞれの言語で主語、後述の位置が異なるなどの語順の差異により起こる場合である。しかし強調箇所が分散することは強調の効果が稀薄になるので、できるだけ近接することが望ましい。また、強調箇所が分散しているとオペレーターがポストエディット時に強調箇所を探し出す場合、第1言語のデータと対応させて第2言語のデータから探し出さなければならず、多くの労力を必要とするという問題点があった。

(発明が解決しようとする課題)

上述したように従来の機械翻訳装置には、第1言語の

データの強調箇所が対応する第2言語のデータに反映されず、後からオペレーターがその箇所を強調する様に修正しなければならなかつた。また、第1言語のデータを強調箇所が第2言語のデータでは分散して訳出されることにより強調の効果が弱まるということが起つた。この様に第1言語のデータの強調箇所が対応する第2言語のデータに反映されないという問題点があつた。

本発明は上述した問題点を解決するためになされたものであり、第1言語のデータの強調箇所に対応する第2言語のデータの箇所を強調する機械翻訳方法及び機械翻訳装置を提供することを目的とする。

【発明の構成】

(課題を解決するための手段)

上記目的を達成するため本発明は、各種情報を入力するための入力手段と、各種情報を出力するための出力手段と、各種情報を記憶する記憶手段を備えた装置上で、原文情報を翻訳処理して訳文情報を得る機械翻訳処理方法であつて、前記入力手段から入力された強調箇所を示す属性データを含む翻訳対象となる原文情報を前記記憶手段に記憶させ、前記記憶手段から読み出した原文情報に強調箇所を示す属性データが含まれているか否かを判定し、原文情報中に強調箇所を示す属性データが含まれていると判定した場合には、翻訳結果である訳文情報を生成する過程で語順に自由度のある部分に対して該属性データが付与されている語句を連続した語順となるよう訳文情報を生成し、前記生成された訳文情報の対応する箇所に強調箇所を示す属性データを反映させて前記記憶手段に記憶させ、前記記憶手段から読み出した訳文情報を前記出力手段に出力させることを特徴とする機械翻訳方法を提供するものである。また、同様に上記目的を達成するため本発明では、強調箇所を示す属性データが含まれた翻訳対象となる原文情報を記憶する原文情報記憶手段と、前記原文情報記憶部から読み出した原文情報に強調箇所を示す属性データが含まれているか否かを判定する属性判定手段と、この属性判定手段で原文情報中に属性データが含まれていると判断した場合には、翻訳結果である訳文情報を生成する過程で語順に自由度のある部分に対して該属性データが付与されている語句を連続した語順となるよう訳文を生成すると共に前記属性データを訳文情報の対応する箇所に反映させる翻訳処理手段とを備することを特徴とした機械翻訳装置を提供するものである。

(実施例)

以下、図面を参照しながら本発明の実施例の説明を行う。

第1図は本発明の一実施例のブロック図を示す。以下、英日翻訳を例にとり説明する。

第1図において、入力部1は翻訳対象となる原文（この場合、英文）の入力や各種コマンドや強調箇所の入力を行う為のもので、通常キーボードが使用される。また

OCR等による入力も考えられる。表示部2は、入力部1より入力した英文や、翻訳部5により訳出された日本文を表示するため、文字や下線、斜文字等が表示でき、例えばCRTディスプレイが使用される。文書記録部3は、第1言語である英文と第2言語である日本文のそれぞれのデータを独立して記録できる他、英文のデータとその翻訳結果としての日本文のデータを対応づけて記録することが可能である。辞書記録部4は、翻訳処理に必要な知識情報として、第1言語である英語の単語・熟語と第2言語である日本語の訳語とを対応づけて記録した語彙辞書などを記録するもので、さらに見出し、品詞、活用形、概念、文法機能などを英語と日本語の両方で記述したものも記録するものである。この他にさらに、翻訳処理する際、第1言語のデータの強調箇所に対応する第2言語のデータの箇所の単語同士を近接して訳出する時に用いる強調ルールが記録されている。翻訳部5は、第1言語のデータである英文を第2言語のデータである日本文に翻訳するものである。これは従来の翻訳部に、英文から他の部分と異なるフォントの斜文字や、下線などや、前記入力部から入力された強調箇所などの強調情報を抽出する機能、前記強調情報を保持する機能、翻訳処理において訳文（この場合、日本文）を生成する際、前記抽出機能により抽出された強調箇所が近接するように制御する機能、翻訳結果に、前記強調情報に基づいて強調処理を行う機能を付加したものである。ここで、強調情報とは、強調箇所の原文での位置情報や斜文字や下線部の表記情報等を含むものである。翻訳制御部6は、本装置の全体を制御するものである。

第2図は、入力部1のキー配列の一例を示す図である。入力部1は英文／和文入力用の文字キー201に加えて、翻訳指示キー202、編集キー203、機能キー204、カーソルキー205及びその他のキーを備えている。

第3図は表示部2の画面レイアウトの一例を示す図である。入力した原文は画面左側の原文表示領域301に表示される。翻訳処理の結果得られた訳文は画面右側の訳文表示領域302に表示される。また、画面上部の編集領域303は各種編集に必要な情報を表示するために用いることができる。

第4図は、翻訳制御部6の動作を説明するためのフローチャートである。翻訳制御部6は翻訳対象文書を指定し、翻訳開始のコマンドが入力部1から入力されることにより動作を開始する。翻訳開始に先立って、まず、原文中に出現する下線、斜文字等の部分的な強調のうち、何をどのように訳文に反映させるかをユーザーに指定させる。最初に、下線部を翻訳結果に反映させるか否かを指定する（ステップ401）。反映させる場合には、反映させる方法を指定する（ステップ402）。通常は原文で下線が付加されている場合、その部分に対応する訳文にも下線を付加するという指定を行う。次に、原文中の斜文字を翻訳結果に反映させるか否かを指定する（ステッ

プ403）。反映させる場合には、反映させる方法を指定する（ステップ404）。通常は原文が斜文字で表記されている場合、その部分に対応する訳文も斜文字で表記する。同様に原文中の太字を翻訳結果に反映させるか否かを指定する（ステップ405）。反映させる場合には、反映させる方法を指定する（ステップ406）。通常は原文が太字で表記されている場合、その部分に対応する訳文も太字で表記する。また、その他のフォントについても同様に翻訳結果に反映させるフォントの種類と、その反映のさせ方を指定する（ステップ407, 408）。通常は原文で使用されているフォントを、その部分に対応する訳文でも同様に使用するという指定を行なう。以上の処理によって、原文中の部分的な強調に対する処理がきまるため、翻訳対象となる文書として第1言語のデータ（データA）を読み込む（ステップ409）。まず、データAの第1文を取り出し（ステップ410）、この中に言語データ以外に強調情報を含んでいるかどうかを検査し（ステップ411）、検査の結果、含んでいる場合には、前記訳文に反映させる方法に従って強調情報を保持する（ステップ412）。なお、この検査方法は後述する。これを翻訳部5へ転送し、翻訳結果を得る。なお、この翻訳過程については後に第5図を用いて説明する。以上の様な処理（ステップ411～413）をデータAの最後の文になるまで繰り返し（ステップ414, 415）、データAの最後の文まで翻訳が終了した場合には、データAとその翻訳結果を文書記録部3に記録して（ステップ416）、処理を終了する。

ここで、言語データ以外に強調情報を含んでいるかどうかを検査する検査方法について述べる。前記入力部1から入力されたデータには、下線や斜文字等の表記情報が付加されている。これは語句の属性として付加されているので、検査対象の語句の属性の有、無しを判別し、さらに属性が有る場合、表記の種類、例えば下線、太字等、を判別する。

第5図は第4図におけるステップ413、すなわち翻訳部の処理の流れを説明するフローチャートである。翻訳部5は翻訳制御部6より翻訳対象となるデータが転送される事により動作を開始する。そのデータが転送されるとまず、ステップ411で言語データ以外に強調情報を含んでいると判断された語句にマークを付ける（ステップ501）。そして、本実施例ではトランスマニア方式を用いているので解析処理（ステップ502）、変換処理（ステップ503）を行い、翻訳結果を生成する前段階のデータを得る。次に、翻訳結果を生成処理する際、ステップ501において同一のマークが付けられた語句に対応する訳語が近接するように語順を決定する（ステップ504）。なお、近接するように語順を決定する処理は第6図を用いて後に説明する。そして、さらに、マークの付いた語句の訳語に、翻訳部5に保持している強調情報を付加する。これはマークの付いた語句から抽出した強調情報を

対応する訳語に属性を付加することにより実現される。以上の処理において原文での強調箇所を対応する訳文の箇所に反映することができる。

第6図は第5図におけるステップ504、すなわち同一のマークを持った語句同士が近接するように語順を決定する処理の流れを説明するフローチャートである。ステップ502, 503で翻訳結果を生成する前段階のデータが得られると、翻訳部5は辞書記録部4に記録している強調ルールにより動作を開始する。その前段階のデータが得られると、まず、そのデータの内、語順の決定している箇所のみ生成を行なう（ステップ601）。これは英日翻訳の場合であれば、日本文における接続詞や述語等の位置であり、日英翻訳の場合であれば、英文における主語や動詞等の位置である。次に、語順に自由度がある語句に、マークが付いた語句が含まれているかどうかを検査し（ステップ602）、検査の結果含まれていなければ語順の決定している箇所にマークが付いていることになる。これでは語順を変更することはできないので従来の方法により翻訳処理を行なう（ステップ607）。なお、上記の検査方法は、ステップ501でマークが付けられたかどうかを検査すればよい。そして、ステップ602でマークが付いた語句が含まれている場合、そのマークが複数であるかどうかを検査し（ステップ603）、複数であればこれらの語句を近接させる（ステップ604）。さらに、語順が決定している箇所にマークが付いた語句があるか検査し（ステップ605）、ある場合にはステップ604で近接させた語句、またはステップ603でマークが付いた語句が1つだけであるとされた語句を、さらに近接させた上で（ステップ606）、従来の方法による翻訳処理を行う（ステップ607）。また、語順が決定している箇所にマークが付いた語句がない場合も同様に従来の方法による翻訳処理を行う。以上の処理において原文での強調箇所に対し対応する訳文の箇所を近接して訳出することができる。

次に、第7図、第8図を用いて、本発明が提供する機械翻訳装置における部分的に強調された原文の翻訳処理の説明を行う。今、第7図（a）のように原文表示領域301に入力した “He told that it was her book.” という文を翻訳する。この時の翻訳結果は第7図（b）の様に訳文表示領域302に「彼はそれは彼女の本だと言つた。」と表示される。次に、第8図（a）のように “He told that it was her book.” という様に “He told” の部分が下線により強調された文を翻訳する。ここで、ステップ401, 402において原文の下線部は翻訳結果に下線として強調するという設定がされているとする。すると、ステップ411の検査の結果、ステップ412で「属性有、下線」という強調情報を保持する。そしてステップ501で “He” と “told” にマークを付け、ステップ502, 503において解析処理、変換処理を行うと、

主語：「彼」（属性有、下線）

述語：「言う」（属性有、過去形、下線）

目的語：「それは彼女の本である」（属性無）

というデータを得る。そして翻訳部5は辞書記録部4に記録している強調ルールに基づいて生成処理を行う。まず、ステップ601で述語の位置が決まるので「言った。」が生成される。ステップ602では “He” と “told” のうち “He” には語順に自由度がある語句なのでステップ603の検査を受ける。その結果 “He” のみの1語なのでステップ605に行き、語順が決定している箇所にマークが付いた語句があるか検査する。その結果、ステップ601で語順が決定している “told” にマークが付いているので、ステップ606で “He” と “told” を近接して「彼は言った。」と訳出される。この後、ステップ607により、「それは彼女の本であると彼は言った。」と訳出される。さらに、ステップ505によって第8図（b）に示す最終的な訳文を得る。

なお、本発明は上述した実施例に限定されるものではない。例えば、上記の実施例においては、下線、斜文字、太字に対し強調処理を行っているが、これらに限定されるものではなく、例えば全文字が大文字である様な単語や波線、反転文字など表記情報を附加させた語句に対する強調等にも適用できる。また、原文の表記を追加、変更しなくても利用者が原文で強調箇所をマークし、それを本装置が保持していればその箇所に対応する訳文の箇所を強調することができる。また、本実施例では原文が下線で強調されている場合、訳文も下線で強調されるが、訳文に反映される方法が色々設定できるので、これに限定されることなく、訳文では波線等の原文とは異なる方法で強調することも可能である。また、本実施例では原文の強調情報を訳文に反映させる方法を1つ1つ設定しているが、これを原文と同一の強調情報を与える場合は一度に設定する様にすることも可能である。

この他に、本実施例では同じマークである下線がついた語句同士を近接させているが、複数のマーク、例えば下線と斜文字等、同士を近接することも可能である。さらに、複数のマークを近接するように処理する際は、どのマークの近接を優先するかを決定する。つまりマーク間で優先順序を設定して近接処理を行うことも可能である。

さらに、原文の強調情報を翻訳部に保持しているが、別に保持部を設けることも容易に可能である。また、本実施例では強調箇所を近接させる処理と強調情報を付加する処理の両方を行っているが、どちらか一方を行っても良く、それを利用者が選択することも可能である。これにより、例えば強調情報を付加する処理のみを用いて、原文の語句が訳文のどの語句に対応しているかを調べることができる。さらに、以上の様な強調処理自体を行なうかどうかの選択を、利用者が決定することも可能である。

9

さらに、この他に、本実施例では、翻訳処理が「解析」「変換」「生成」の手順で行われるトランスファー方式に基づいて述べているが、ピボット方式による翻訳処理でも可能である。これは、本発明は「生成」の段階で強調処理を行うものであるので、翻訳処理の方式に影響されるものではないからである。

要するに、本発明はその主旨を逸脱しない範囲で種々変更して実施することができる。

〔発明の効果〕

以上述べたように本発明によれば、機械翻訳装置において、原文である第1言語の文章中に斜文字や下線等により部分的に強調されている語句や利用者が指定した強調箇所がある場合に、その強調情報を保持する。そして強調された語句を強調ルールを用いて第1言語のデータの強調箇所に対応する第2言語のデータの箇所の単語同士を近接して訳出したり、保持している強調情報に従い強調された語句の訳語に、原文に成されたものと同様の強調効果を付加する。これらによって、原文の強調箇所を訳文に反映させる事ができる。従ってオペレーターの*

10

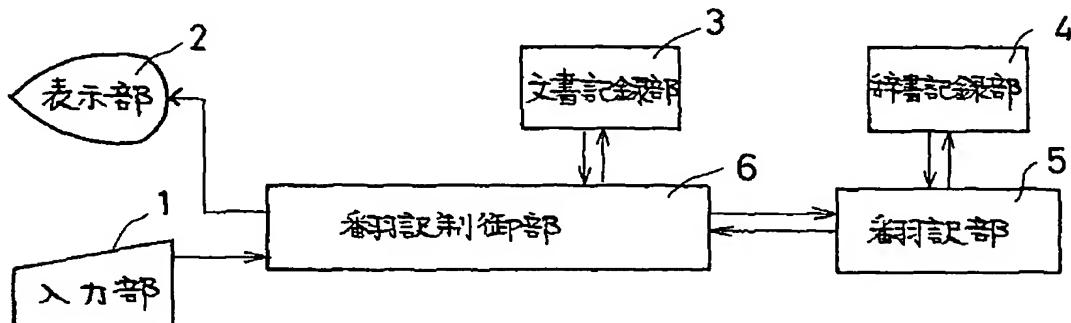
*負担を軽減する上に、原文の意図が理解しやすい文章を得ることができる。

【図面の簡単な説明】

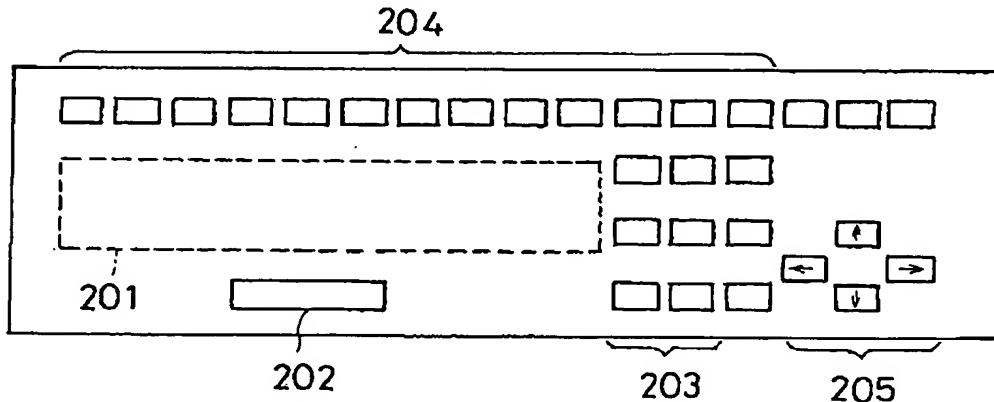
第1図は本発明の一実施例の全体のブロック図、第2図は第1図に示した入力部のキー配列の例を示す図、第3図は第1図に示した表示部のレイアウトの例を示す図、第4図は本発明の一実施例で用いる翻訳制御部の処理の流れの一例を示すフローチャート、第5図は第4図における翻訳処理の詳細を示すフローチャート、第6図は第5図における近接するように語順を決定する処理の詳細を示すフローチャート、第7図から第8図は本発明の一実施例における処理の説明の為のデータの例を示す図である。

- 1 ……入力部
- 2 ……表示部
- 3 ……文書記録部
- 4 ……辞書記録部
- 5 ……翻訳部
- 6 ……翻訳制御部

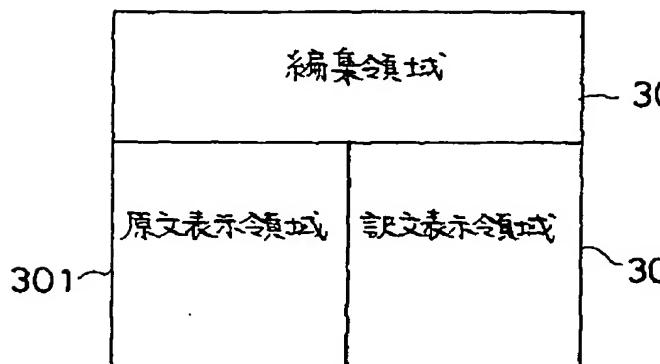
【第1図】



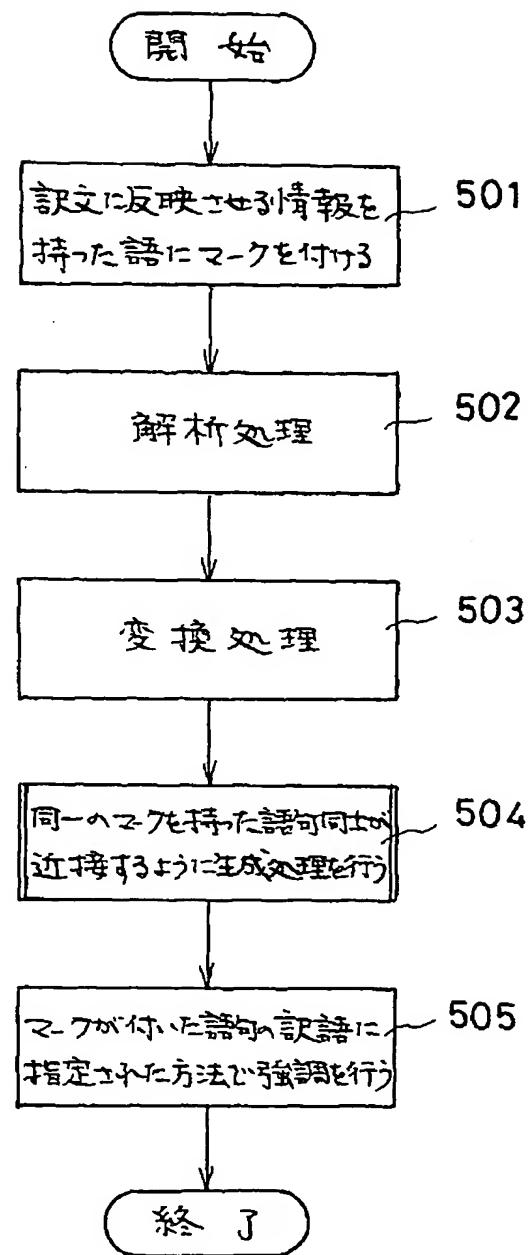
【第2図】



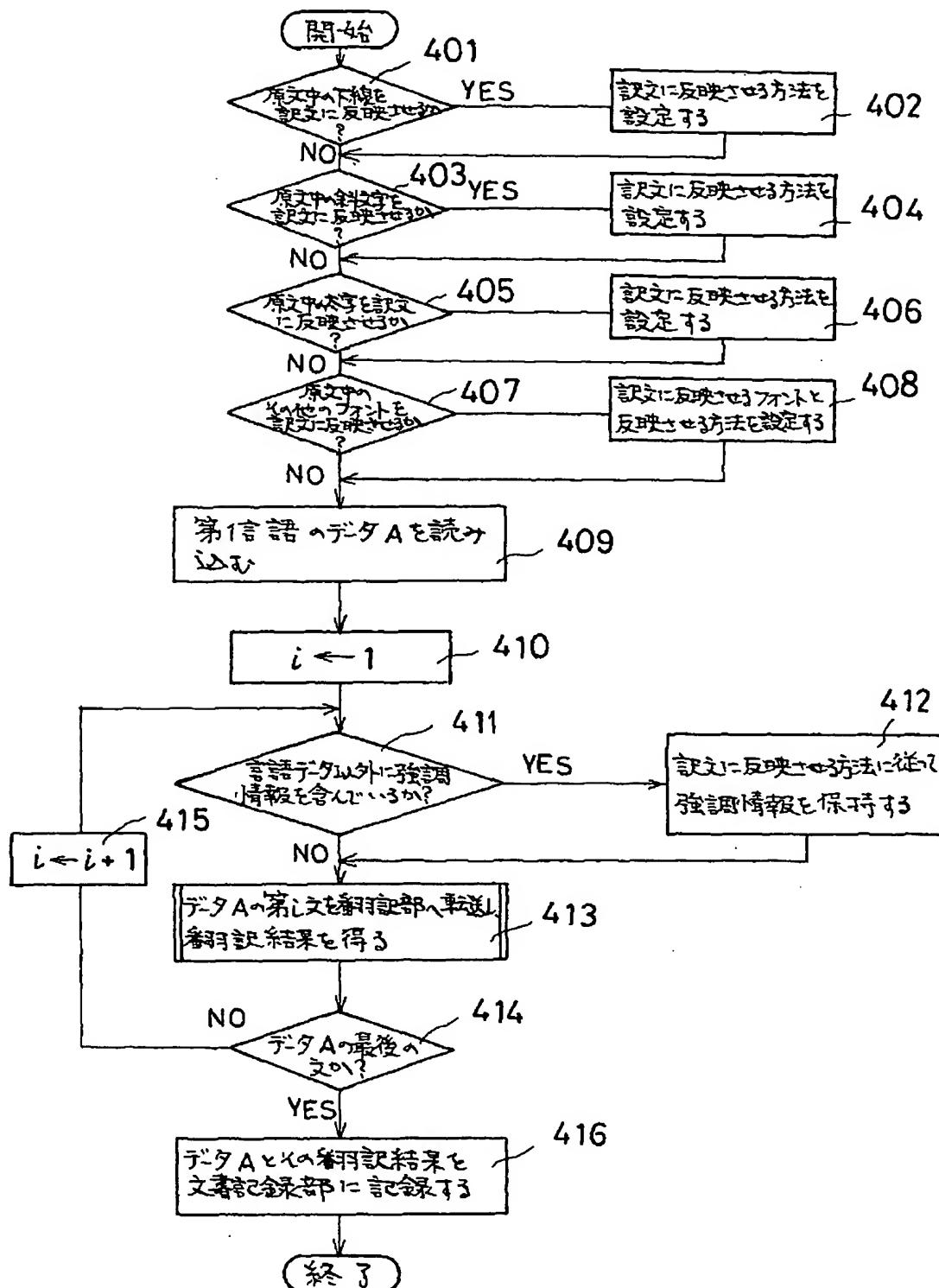
【第3図】



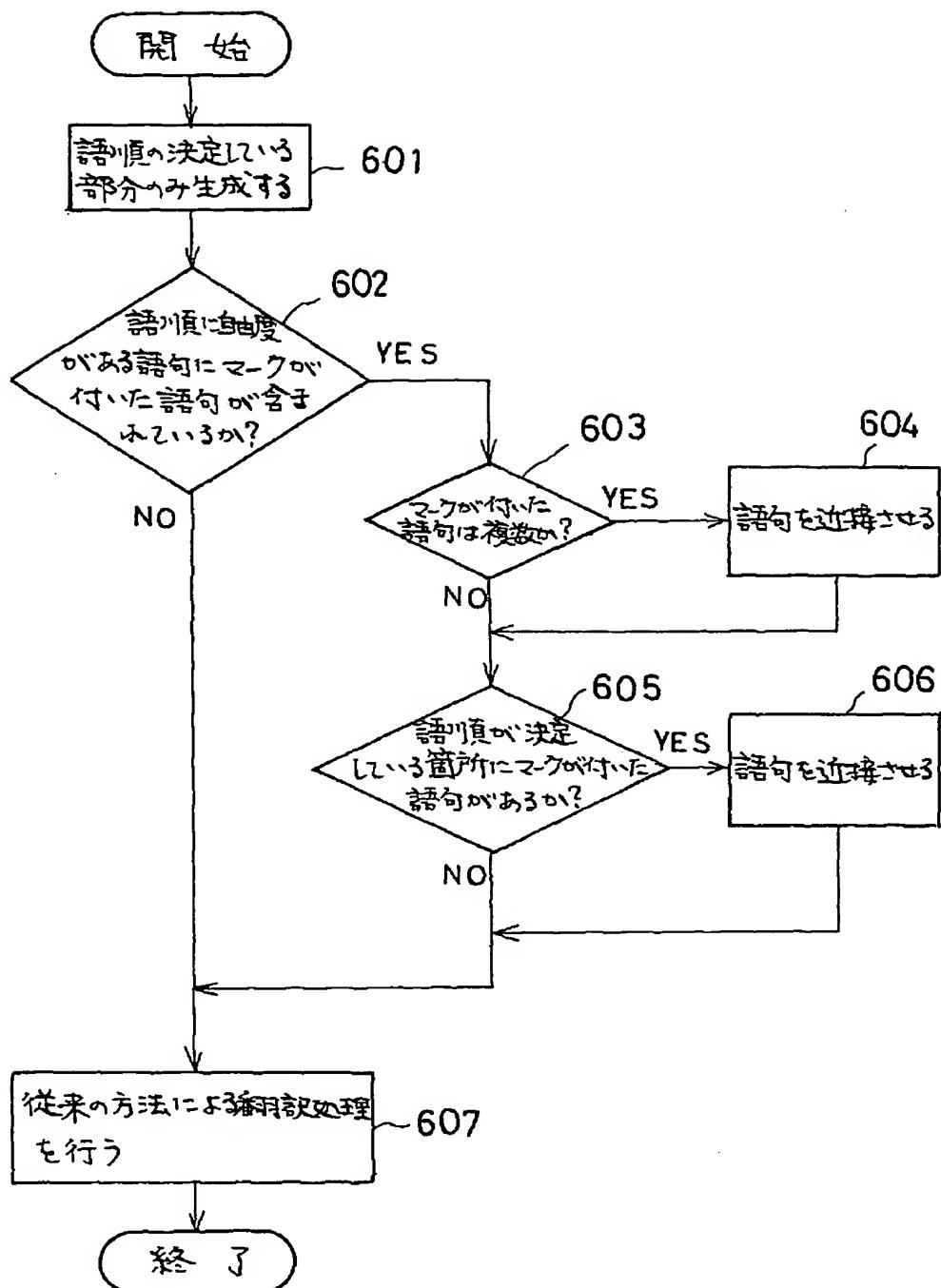
【第5図】



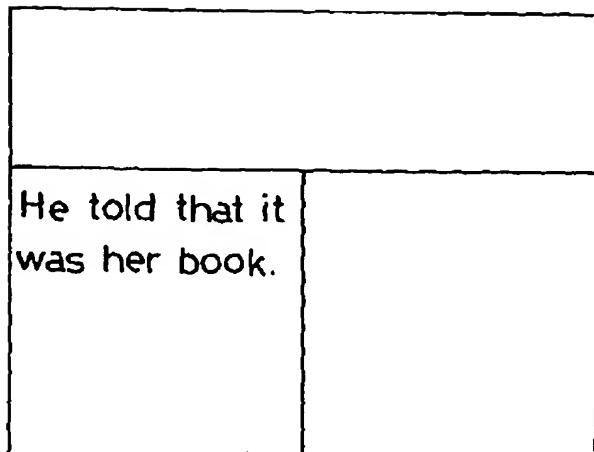
【第4図】



【第6図】

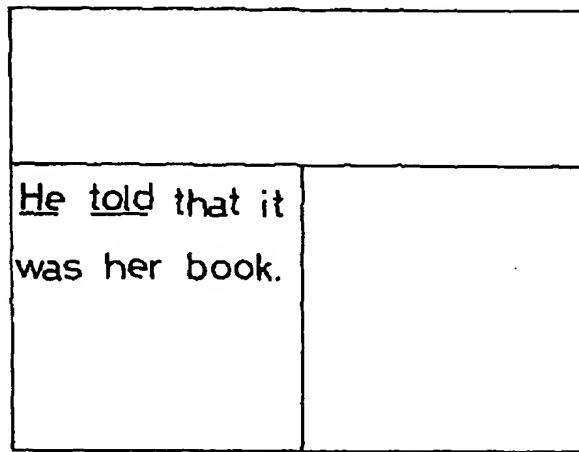


【第7図】

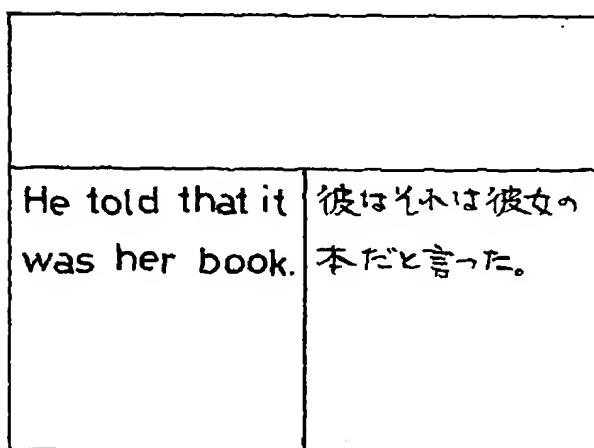


(a)

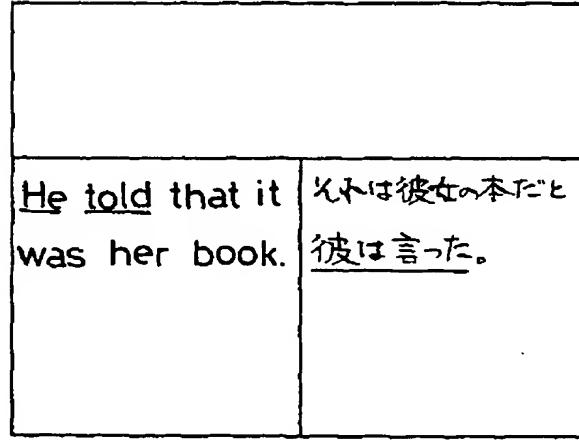
【第8図】



(a)



(b)



(b)

フロントページの続き

(72)発明者 武田 公人
 神奈川県川崎市幸区小向東芝町1番地
 株式会社東芝総合研究所内

(56)参考文献 特開 昭62-44873 (JP, A)
特開 昭63-220361 (JP, A)
特開 平1-78372 (JP, A)

(58)調査した分野(Int.Cl.7, DB名)

G06F 17/20 - 17/28